Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 7, July 2020, pg.97 – 108

EFFICIENTLY ANALYZING AND DETECTING FAKE REVIEWS THROUGH OPINION MINING

Ashwini.M.C; Padma.M.C

Abstract- Recently, Sentiment Analysis (SA) has become one of the most interesting topics in text analysis, due to its promising commercial benefits. One of the main issues facing SA is how to extract emotions inside the opinion, and how to detect fake positive reviews and fake negative reviews from opinion reviews. Moreover, the opinion reviews obtained from users can be classified into positive or negative reviews, which can be used by a consumer to select a product. The growth of e-commerce businesses has attracted many consumers, because they offer a range of products at competitive prices. One thing most purchaser rely on when they purchase online is for product reviews to conclude their decision. Many sellers use the decision to impact the review to hire the paid review authors. These paid review authors target the particular brand, store or product and write reviews to promote or demote them according to the requirements of their hired employees. This paper aims to classify Amazon product reviews into groups of positive or negative polarity by using machine learning algorithms. In this paper, we analyze online product reviews using SA methods in order to detect fake reviews. SA and text classification methods are applied to a dataset of product reviews. This paper focuses on detecting fake reviews from a set of product reviews by simulating fake reviews that incorporates various types of opinion spam review features and building a training set and then classifying it using Naïve Bayes classification and ensemble classification model like random forest to test the accuracy of the model.

Keywords- Sentiment Analysis; Fake Reviews; Naïve Bayes; Support Vector Machine; k-Nearest Neighbor; random forest.

INTRODUCTION

Sentiment Analysis (SA) is one of the extensive domains of machine learning. Opinion Mining (OM), also known as Sentiment Analysis (SA), is the area of study that analyzes people's opinions, evaluations, sentiments, attitudes, appraisals, and emotions towards entities such as services, individuals, issues, topics, and their attributes [1].

There are three major classification in SA namely: document level, Sentence level, aspect level or phrase level of an analysis process that will determine the different task of SA.

Document level

Document level sentiment classification executed on the overall sentiments expressed by authors. Documents classified according to the sentiments instead of topic. It is to summarize the whole document as positive or negative polarity about any object (mobile, car, movie, and politician etc.).

Sentence level

Sentence level sentiment classification models extract the sentences contains opinionated terms, opinion holder and opinionated object. It is one level deep to document level and just concerns to the opinionated words but not the features. Number of positive and negative words counted from sentences if positive words are maximum then opinion about object is positive and if the negative words are more than opinion is negative otherwise neutral.

Aspect or Phrase level

Opinion Mining The phrase level sentiment classification is a much more Pinpointed approach to opinion mining. The phrases that contain opinion words are found out and a phrase level classification is done. But in some other cases, where contextual polarity also matters, the result may not be fully accurate. Negation of words can occur locally. But if there are sentences with negating words which are far apart from the opinion words, phrase level analysis is not desirable. The process is Identifying Opinion Words, the role of negation words and Clauses.

Our work is largely focused to SA at the document level, more specifically on Amazon product reviews dataset.

Nowadays, customers prefer buying most products or services through e-commerce or online portals have given rise to new techniques for marketing as well as influencing customers decision i.e. reviews. Reviews present a new way to learn about customer preferences, product quality as

well as product short comings. opinion sharing on a product/service is based on their personal experience which is called as reviews. Sometimes, time is more precious than money, therefore, reading about product reviews before buying the product becomes a habit, especially for potential customers. If the customer ready to buy a product, they usually read reviews of other customers about the current product. The reviews play an important role in how the end users view a product. It is human nature to judge a metaphorical book by its reviews. No matter how good a user experience with it is, they tend to believe in the word of mouth. Thus, fake reviews are a hidden threat to e-commerce businesses. It is unethical but widely practiced. The advancement in the technology has brought with it tools to deal with said fake reviews and the subsequent fallout. using these tools, e-commerce sites can curb this malpractice and bring integrity to the e-commerce business.

Some of the typical characteristics of fake reviews are [2],

1. Less Information about the Reviewer: Users who have fewer social connections and who do not have profile information are usually impostors and are likely to post only a few reviews which are fake reviews.
2. Review Content Similarity: Spammers often copy their own reviews or reviews of other users and often write duplicate or near duplicate reviews. These reviews may indicate spam reviews.
3. Short Reviews: As spammers are interested in making quick profits, they tend to write very short reviews with a lot of grammatical errors and excessively use capitals, numerals and all capital words. They also focus on brand names of a product.
4. Sudden uploading of reviews in the same timeframe: It is found that one of the best ways to detect fake reviews is by looking at the timestamp of the reviews and if a batch of reviews is uploaded at the same time then this is a spam indicator.
5. Focus on personal information: Genuine reviews usually focus on spatial information but spam reviews focus on irrelevant personal information.
6. Excessive use of positive and negative words: Spammers often use a lot of positive and negative words in a review which might not be the necessity in the context of the review. Further they also write reviews based on the product description and not on their experience of using the product.

Fake reviews have become one of the major concerns and a big threat in today's world. SA methods and machine learning techniques are expected to have a major positive effect, especially

for the detection processes of fake reviews in Amazon product reviews, e-commerce, social commerce environment and other domains.

In this paper, fake review detection has been considered as binary classification problem with the two classes being: fake and genuine. This paper focuses on detecting fake reviews from a set of product reviews by simulating fake reviews that incorporates various types of opinion spam review features and building a training set and then classifying it using Naïve Bayes classification and ensemble classification model like random forest to test the accuracy of the model.

The remainder of this paper is structured as follows: the next section discusses the work done in the fake review detection domain. Section 3 gives the methodology. Section IV explains the experiment results, v. comparison between different classifiers and finally, Section VI presents the conclusion and future works.

Related Work

A. Detecting spam review through sentiment analysis

In paper [3] the author detects spam reviews by incorporating the concept of sentiment analysis.

CONTRIBUTION:

The author first creates a sentiment lexicon which combines the data present in existing sentiment lexicons such as Senti WordNet and MPQ, along with designing sentiment lexicon specifically for products. Further he calculates sentiment score which is the sentiment polarity of a review. Also, he calculates other parameters such as sentiment ratio (ratio of sentiment sentence to all sentences) and difference of sentiment polarity (inconsistency in sentiment score and rating score). Next, he constructs discriminative rules to classify the reviews as spam. Finally, he combines the discriminative rules with the time series method to detect spam in a spam detection algorithm.

OBSERVATION:

Here, the reviews are pre-processed and ordered by time. For each subset of reviews, each review is checked to find whether it satisfies any of the discriminative rules within the given time window. If the result is positive then the store view is categorized as spam. The proposed sentiment score method outperforms the rating and word counting methods with an accuracy of 85.7%. The discriminative rules here are derived based only on three parameters namely sentiment score, sentiment ratio and discrepancy between rating and sentiment. However, the author has not considered the effect of behavioral parameters such as Individual rating deviation, Individual

content similarity and Individual early time frame for deriving discriminative rules and this is one of the areas which needs to be explored.

B. Fraud Detection in Online Reviews by Network Effects

In this paper [4] the author has proposed a model which uses network classification to detect frauds. algorithm called Signed Inference Algorithm is used which is a message passing algorithm. Here a set of messages are exchanged between the users and products in an iterative fashion until the messages stabilize. Finally, the marginal probabilities are calculated and final belief of user having a particular label is computed. The class labels of both users and products are inferred by the final belief vectors. The marginal class probabilities of products, users and reviews are used to order each set of items in a ranked list.

OBSERVATION:

The proposed method is very simple and does not require labelled data for classifying reviews. However, the author has considered only user ratings as the basis to detect spam and has not considered the sentiment analysis of text found in reviews.

C. Detecting Fake Reviews by the principle of Collective Positive Unlabeled Learning

In this paper [5] the author has proposed a model of detecting fake reviews by learning from positive and unlabeled examples.

CONTRIBUTION:

Here a heterogeneous network consisting of users, reviews and IP addresses is being considered for the learning model where a classification algorithm called MHCC (Multi- typed Heterogeneous Collective Classification) algorithm is used first and then this is extended to Collective Positive and Unlabeled Learning. The MHCC algorithm considers a heterogeneous network of users IP addresses and reviews as input along with feature matrix of reviews, user sand IP addresses.

The class label of the review is the desired output. The algorithm has both initialization and prediction steps where the adjacency matrix is computed and then the classifier is trained. The initial classifier gives a rough estimate of the review being in fake review class and the labels of IP and users are derived from majority class labels of their related reviews.

OBSERVATION:

In the prediction step a relation feature matrix is constructed from the estimate labels of the neighboring nodes and this is used as the basis to train3 different classifiers for reviews, users and IPs which will provide more accurate results.

However, this algorithm treats all unlabeled samples as negative and the adhoc labels of users and IPs may not be accurate as they are derived from labels of neighboring reviews. Hence an extended model called Collective Positive Unlabeled model is used where new labels are updated with respect to confident positives and negatives from all entity types. This algorithm allows initial labels to be violated if current probability estimate indicates opposite prediction.

The model outperforms baseline algorithms like Logistic Regression and also detects a large number of potential fake reviews hidden in an unlabeled set thus improving the training data.

Heterogeneous network can be extended to include metadata information of users such as timeframe of writing reviews, number of written reviews and demography of the reviewer. This can be an active area to be explored.

METHODOLOGY

To accomplish our goal, we analyze a dataset of Amazon reviews using the Weka tool for text classification [6]. In the proposed methodology, as shown in Figure 1, we follow some steps that are involved in SA using the approaches described below.

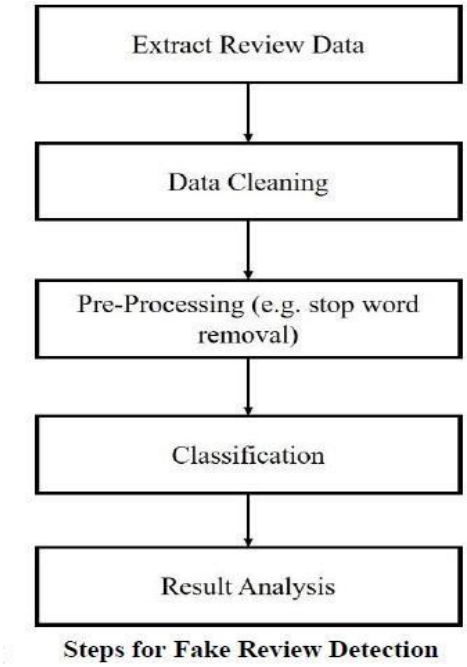


Fig1.Steps for Fake Review Detection

3.1. DATA COLLECTION:

To provide an exhaustive study of machine learning algorithms, the experiment is based on analyzing the sentiment value of the standard dataset. We have used the original dataset of the Amazon product reviews to test our methods of reviews classification. Data extracted from dataset is used as the unlabeled data, labeled dataset created by Ottertail is used for both training and testing purpose in this method. As the dataset created already preprocessed so we only do some preprocessing on the unlabeled dataset. We use reviews extracted from as the unlabeled data in this learning method.

3.2 CLEANING OF DATA:

The Deceptive fake review dataset contains no cleaning of data because it is in the presentable form. The Amazon product dataset requires data cleaning because the rows in metadata and content file are not equal. The proposed technique cleans the data by removing rows from metadata that are not in content files. In this way, system has the same number of rows in both the files and then the data is combined. Now, the system is ready for further processing.

3.3 LOADING THE DATA:

The dataset is in comma separated values (CSV) format so Pandas library of python is used to load it into the desired python Integrated Development Environment (IDE). Amazon dataset contains unequal distribution of deceptive and truthful reviews so the system is loaded with only the 110,580 rows data in which 36,860 reviews are deceptive and remaining are genuine reviews. To balance the dataset system framework, duplicate the deceptive reviews so the system contains the equal 147,440 reviews half of them are deceptive and half are genuine reviews. In this way, overfitting is avoided and model is trained accurately.

3.4. VISUALIZATION OF DATA

After visualizing, the data system finds out that the length of fake reviews is long and contains words that are more positive, more punctuation, and repetition of words.

3.5. SPLITTING THE DATA

For splitting the data into training and testing one, proposed technique uses the most common form of splitting that is 20% for testing and 80% for training.

3.6 DATA PRE-PROCESSING:

Unstructured data in MS Excel format acquired from the source is converted into structured data i.e. in My SQL Database format. Data preprocessing plays a significant role in many supervised learning algorithms. We divided data preprocessing as follows:

- **StringToWordVector**

To prepare the dataset for learning involves transforming the data by using the StringToWordVector filter, which is the main tool for text analysis in Weka. The StringToWordVector filter makes the attribute value in the transformed datasets Positive or Negative for all single-words, depending on whether the word appears in the document or not. This filtration process is used for configuring the different steps of the term extraction. The filtration process comprises the following two sub-processes:

• Configure the tokenizer

This sub-process makes the provided document classifiable by converting the content into a set of features using machine learning.

• Specify a stopwords list

The stopwords are the words we want to filter out, eliminate, before training the classifier. Some of those words are commonly used (e.g., "a," "the," "of," "I," "you," "it," "and") but do not give any substantial information to our labeling scheme, but instead they introduce confusion to our classifier. Stopwords removal helps to reduce the memory requirements while classifying the reviews.

- **Attribute Selection**

Removing the poorly describing attributes can significantly increase the classification accuracy, in order to maintain a better classification accuracy, because not all attributes are relevant to the classification work, and the irrelevant attributes can decrease the performance of the used analysis algorithms, an attribute selection scheme was used for training the classifier.

- **Feature Selection**

Feature selection is an approach which is used to identify a subset of features which are mostly related to the target model, and the goal of feature selection is to increase the level of accuracy. Irrelevant data can reduce the performance and accuracy of the classifier. Thus, it is better to eliminate the irrelevant data before extracting the features. System is studied by two feature

selection methods: Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) for the selection of features related to our dataset. In the proposed approach, system uses Count Vectorizer that converts each review into bag of words and is used to tokenize the set of words described in the reviews and after its system applies TF-IDF transformer. The model converts the collection of text documents into a matrix of token counts. It is 2-dimensional matrix where 1-dimension represents the vocabulary and other dimension of actual document as described in Table 1.

Table1. Count vectorizer overview

	Word 1	Word 2	.....	Word N
Review 1	0	2	.....	1
Review 2	0	1	.....	1
.....	1	0	.....	2
Review N	2	1	.....	0

Since, there are a lot of zero involved in this matrix so it is called sparse matrix. The TF-IDF Transformer is the weighted metric used in text mining and is used to measure how important is the word in that dataset. Importance of the word increases based on how many times the word appears in the dataset. Each word is assigned a respective TF-IDF score. For a word t in document d, the weight W (d, t) of the word t in the document d is given as describe in Eq. 1:

$$W(d, t) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right) \quad \textbf{..... (1)}$$

Where, TF (t, d) is the number of occurrences of word t in document N is the total number of documents (reviews) in the dataset DF (t) is the number of documents (reviews) containing the word t.

3.7. Classification Process

Fig. 2 shows the classification process of Fake review detection that how the proposed system is working to classify the reviews into genuine and fake ones. It starts with collecting the data, the next step is to preprocess the data including removing punctuation and stop words from the text of the reviews, then system converts the text into lower case, Attribute selection and feature selection and the last step of the preprocessing the system. After preprocessing, extract features using Count Vectorizer that converts each review into 2-D matrix and then apply the TF-IDF transformer that gives weight to each word. After the feature selection, the last step in the classification process is

to train the classifier. The proposed architecture is tested by applying three different supervised machine learning algorithms including SVM, Naïve Bayes, Logistic Regression.

3.8. Predictions and Evaluation

After successfully training the data, system is applied with the testing data to predict unseen data in order to find out whether it is deceptive or genuine. On the Deceptive Opinion Spam dataset of product reviews, system has achieved the accuracy of 90%. Classification process of reviews into truth and deceptive ones is show in Fig.2.

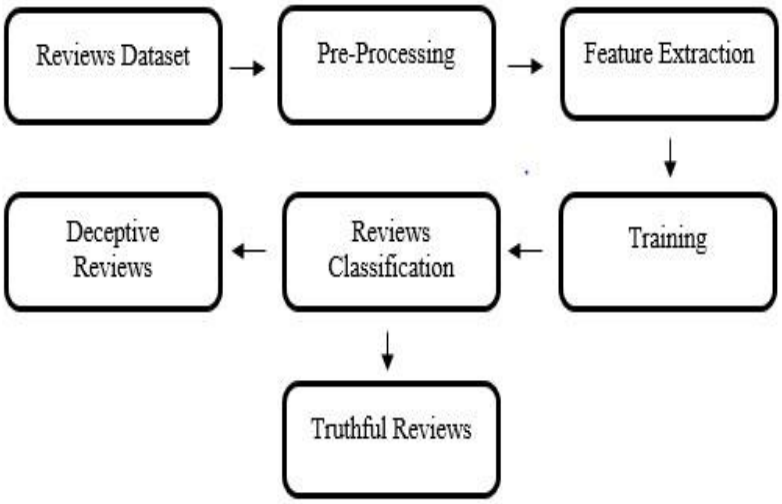


Fig. 2. Fake Review Detection Classification Process

RESULTS AND VISUALIZATION

Amazon Dataset: The proposed architecture uses the SVM classifier to train the model that consists of 147,440 rows. Confusion matrix of this dataset with and without Normalization is given in Table 2 and Table 3.

Table 2. Confusion Matrix of Amazon reviews dataset

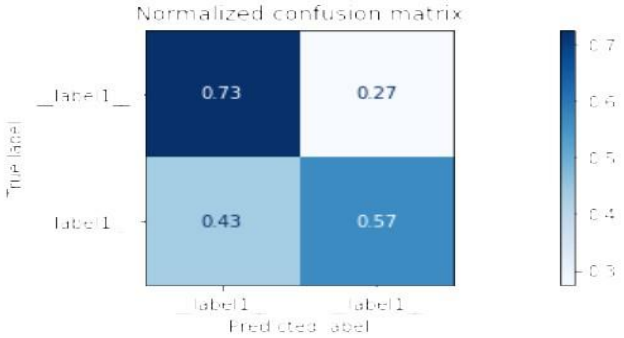
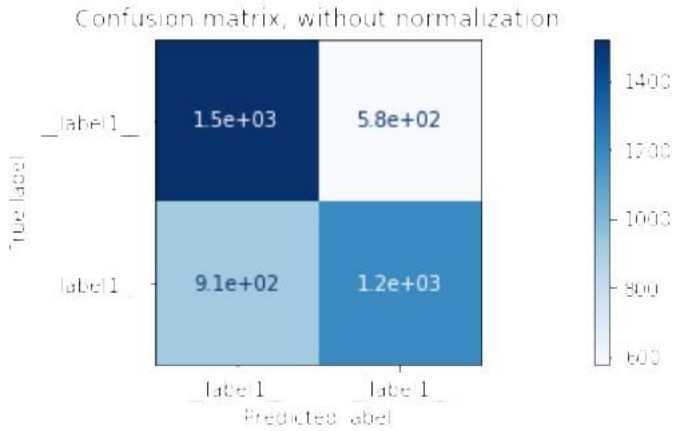


Table 3. Confusion Matrix without normalization

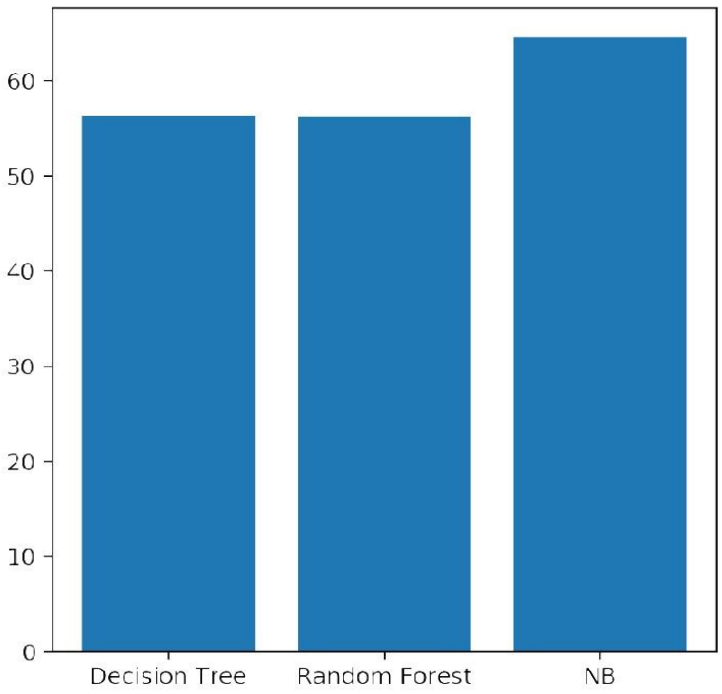


COMPARISON BETWEEN DIFFERENT CLASSIFIERS

The proposed system is tested with three different classifiers to find the best one and it is observed that SVM performs best as compared to the others three classifiers.

Amazon Reviews. Amazon reviews dataset consists of 147,440 reviews. 80% data is used for training and remaining 20% for testing. Performance on different classifiers is shown in Table 4.

Table 4. Comparison of classifiers for Amazon Reviews



CONCLUSION AND FUTURE WORK

In this way we proposing system will automatically classify user opinions into fake or genuine. This automatic system can be useful to customers as well as business organization. Business organization can monitor their product selling by analyzing and understand what the customers are saying about products. Customers can make decision whether he/she should purchase or not purchase the products. This can helpful to people to buy valuable product and spend their money on quality products. Thus, this system makes e-commerce trustworthy.

Future work includes collecting live review data from different review websites. Computer aided generation of fake reviews while incorporating the different features and context aware classification to avoid misclassification of fake reviews. Content aware classification is necessary as it helps to identify sarcasm and other human emotions that is missed in the given model.

REFERENCES

[1]. Elshrif Elmurngi, Abdelouahed Gherbi," Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques." data analytics, vol.11 no 1 & 2, 2018.

[2]. A. Lakshmi Holla, Kavitha K.S, "A comparative Study on Fake Review Detection Techniques." IFERP vol 5, Issue 4, April 2018.pp.641-645.

[3]. Q. Peng and M. Zhong, *"Detecting spam review through sentiment analysis."* JSW, vol. 9, no. 8, pp.2065–2072, 2014.

[4]. L. Akoglu, R. Chandy, and C. Faloutsos, *"Opinion fraud detection in online reviews by network effects."* ICWSM, vol. 13, pp. 2–11, 2013.

[5]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, *"Spotting fake reviews via collective positive-unlabelled learning,"* in Data Mining (ICDM), 2014 IEEE International Conference on. IEEE, pp. 899–904, 2014.

[6]. Ata- ur- Rehman, Nazir M. Danish, "Intelligent Interface for Fake Product Review monitoring and Removal", 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) Mexico City, Mexico. September 11-13, 2019.

[7]. H. Sun, A. Morales, and X. Yan, *"Synthetic review spamming and de- fense"* ,in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1088–1096, 2013.

[8]. L. Akoglu, R. Chandy, and C. Faloutsos, *"Opinion fraud detection in online reviews by network effects."* ICWSM, vol. 13, pp. 2–11, 2013.

[9]. S. Zhao, Z. Xu, L. Liu, and M. Guo, *"Towards accurate deceptive opinion spam detection based on word order-preserving cnn"*, arXiv preprintarXiv:1711.09181, 2017.

[10]. V. Sandules cu and M. Ester, *"Detecting singleton review spammers using semantic similarity"*, in Proceedings of the 24th international conference on World wide web. ACM, 2015, pp.971-976.

[11]. A. Lakshmi Holla , Dr Kavitha K.S "A Comparative study on fake review Detection Techniques", IJERCSE, vol.5, issue.4, pp.641-645, 2018.

